Bayesian data mining, with application to benchmarking and credit scoring

PAOLO GIUDICI

Department of Economics and Quantitative Methods, University of Pavia,

Via San Felice n. 5, 27100 Pavia, Italy

giudici@unipv.it

Summary

The purpose of this article is to show that Bayesian methods, coupled with Markov Chain Monte Carlo computational techniques, can be successfully employed in the analysis of highly dimensional complex datasets, such as those occurring in data mining applications. Our methodology employs conditional independence graphs to localise model specification and inferences, thus allowing a considerable gain in flexibility of modelling and efficiency of the computations.

Some key words: Bayesian model selection; Credit Scoring; Financial Benchmarking; Graphical Models; Markov Chain Monte Carlo Methods.

1 Introduction

The purpose of this article is to show that computational Bayesian methods can be successfully employed in the analysis of data mining applications. In particular, we use graphical models to localise model specification and inferences, thus allowing a considerable gain in flexibility of modelling, efficiency of the computations and interpretability of the inferential results. Furthermore, by employing Markov Chain Monte Carlo methods we provide a simple and efficient way of calculating model scores which allows to perform model selection on the space of all possible decomposable graphical models that describe the association structure of the data at hand.

The latter turns out to be quite useful in data mining contexts, where a priori subject matter information is not sufficient to restrict attention to a limited number of models.

To illustrate our methodology we shall consider two applications of current attention in data mining: financial benchmarking and credit scoring. Both constitute real and challenging applications with which to test our proposed Bayesian scoring method. Although they can indeed be analyzed by simpler, and more traditional methods, our proposed methodology allows to extract further information, in the form of conditional independence structures, and associated probability scores, that may be very valuable in a data mining context, where the purpose is mainly exploratory.

The applications will be presented in the beginning of the paper, in Section 2, to emphasize that the methodology is indeed driven by the problems to be solved.

Section 3 is dedicated to a brief review of the methodology that will be employed to analyze the applications, namely, Bayesian analysis of graphical models and Markov Chain Monte Carlo model selection. In Section 4 we shall present the actual application of our methodology and, finally, Section 5 contains some concluding remarks.

2 The considered applications

2.1 Benchmarking for investment funds

In the last few years investment funds have played an important part in the investment choices of savers. The portfolio composition and the extra-return (with respect to a risk-free baseline) are two factors of primary importance in such choice.

Our objective here is to study the determinants of the extra-return. In particular, we seek to understand the relationships of the extra-return with the financial market indices, often used as benchmark predictors of the return itself.

The problem is that, with the current rapid evolution of financial markets, very little is known on the association structure between the return and the available benchmarks. In order to predict the fund return, financial analyst typically consider, for simplicity, one benchmark index. The accuracy of the resulting predictions depends very much on the explanatory power of the chosen benchmark. However, especially in periods of rapidly changing markets, a multivariate set of indices is necessary, but the choice of which benchmarks to choose is quite difficult.

This is why model selection in this case represents a challenging data mining problem. A good statistical procedure should consider as many alternative models as possible, and should take into proper account not only direct relationships between the return, the benchmarks, and the other available variables, but also their indirect relationships, as the multicollinearities between the predictors are typically very high.

In order to address the above model selection problem, we have considered data kindly supplied by an Italian investment fund company, for the fund named ARCA RR. We have studied the period: February 1992 - January 1998 with observations collected by the end of each month.

The available variables are thirteen, and concern: the quotation, which gives the return of the investment fund, the portfolio composition (namely: percentages of liquidity, Italian treasury bonds, Italian bonds, foreign bonds, convertible bonds and shares), the net collection and also the principal financial market indices, which are used in the Italian financial market as benchmarks, namely: the BTP index, the CCT index, the CTE index, the CTO index, the General index, the J.P.MORGAN index.

In the analysis we have defined the extra-return of the fund as the additional return realised with reference to the risk-free rate of return (in this case the BOT index, composed of short-termed Italian treasury bonds). Consequently, we have treated also the performances of the market indices as differences from the BOT index.

At first we have considered two linear models based on the financial theories known as Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT) (for an introduction on this matter see for instance Elton and Gruber, 1995, and Francis, 1991).

The CAPM is a very simple linear model in which the only relevant factor to determine the extra return is the benchmark; clearly it is very important to establish a priori with great precision which factor to use. Here we shall assume, as a benchmark, what was adopted as such by the investment fund company in the considered period.

The APT model is a less parsimonious linear model and, as the CAPM, it is based on a priori economical hypotheses. When such hypotheses are not firm, the model is built placing links from the extra-return to the explanatory variables that have the highest correlation values. In other words, the model is built exploratorily considering marginal correlations, without considering indirect relations among the variables, as we shall instead do when using graphical models.

In order to select one of the many possible APT models, we have used stepwise backward model selection on all the thirteen available variables. The statistical performance of the models has been compared using both measures of goodness - of - fit (such as F-tests and multiple R-squared) and measures of predictive performance (such as cross-validation errors).

We now present the estimated regression lines corresponding to the final selected models.

The estimated CAPM model is, when choosing, as a benchmark, *Genind*, the General index of the Bank of Italy:

$$EXTRARETURN = 0.8882 \times Genind.$$

On the other hand, the final APT model we have obtained is expressed by the following multiple regression model:

$$
\begin{aligned}
EXTRARETURN &= 2.1049 \times BTPind - 1.91511 \times Genind \\
&+ 0.9153 \times CCTind,
\end{aligned}
$$

where $BTPind$ and $CCTind$ are two indices, mostly composed, respectively, of BTP and CCT, which are medium or long-termed Italian treasury bonds.

2.2 Credit scoring

Credit scoring (see for instance Hand and Henley, 1997) is a class of statistical methods employed to classify creditors in two risk categories: "good" and "bad" payers. By credit risk we mean the probability of a delay in the repayment of the credit granted. In the case of a delay, the creditors will be said to be not credit reliable.

In order to estimate such probability, it is necessary to construct a statistical model that adequately summarizes a large database of information that may be related to the credit behaviour of the individuals. For instance, in a bank such information may be available in the form of a set of demographic information as well as from the operations registered on each individual's account.

Credit behaviour can be influenced by many behavioural aspects; the variables available in the database may capture some of them, but always in a highly interdependent way and, typically, in a way often difficult to establish a priori.

This is why statistical credit scoring constitutes, especially with the recent availability of large databases, an important data mining problem where the use of graphical models, to consider correctly indirect dependencies between variables, and of

Bayesian model selection methods, to take into account model uncertainty, may be key factors.

In fact, Hand *et al.* (1997) have already proposed to employ (non Bayesian) graphical models for credit scoring. In this paper we shall show that a Bayesian approach is also feasible, and is indeed more suited in the presence of strong model uncertainty, as it occurs in data mining contexts.

The dataset we have considered to evaluate our method consists of 1000 observations on creditors of a southern German bank, for which 21 variables are available. The data can be downloaded from the web page of the Institute of Statistics at the University of Munich: http://www.stat.uni-muenchen.de/data-sets/credit.

Given the extremely high sparseness of the data, we have performed a preliminary screening of the variables, following Fahrmeir and Hamerle (1994). We have thus obtained the following nine binary random variables:

(X_1) Gender

(X_2) Marital status: single, non single

(X_3) Banking account ?

(X_4) Good history of banking account ?

(X_5) Good repayment of past credits ?

(X_6) Large amount of the given credit ?

(X_7) Use of the credit: private, professional

(X_8) Credit deadline: short or long term

(X_9) Credit reliability ?

Note that the data are stratified: in the sample, 700 individuals are credit reliable and 300 are not credit reliable.

Although a simple statistical analysis of this dataset may be straightforward, in order to understand the association structure between the behavioural variables we have to consider, in the absence of a priori information, 2^{36} possible graphical models. Choosing one model alone would lead to underestimate model uncertainty.

However, for comparison purposes, a classical backward procedure, with a significance level of 5%, on all 9 variables, leads to the following results:

a) Credit reliability is conditionally independent on gender.

b) Credit reliability is conditionally independent on the amount of the given credit.

c) Credit reliability is conditionally independent on having an account, but not on having a good account.

d) Credit deadline seems to be the variable which is mostly related to the others.

3 Bayesian analysis of graphical models

We now briefly review the methodology we propose to perform Bayesian data mining. We shall only recall the main aspects. More details are contained in Giudici and Green (1999), for continuous random variables, and Giudici, Green and Tarantola (1999), for discrete random variables. For an introduction to graphical models see Whittaker (1990) or Lauritzen (1996).

Let a graph g be described by the pair (V, E), where V is a set of vertices, each of which corresponding to a random variable, and E is the set of edges between such vertices. We shall consider here only undirected edges, such that, if $(a, b) \in E$ also $(b, a) \in E$.

A graphical model is a probability distribution which is Markov with respect to a graph g. Denote it with P_g.

Two important classes of graphical models are (pure) continuous and (pure) discrete graphical models, containing only, respectively, continuous and discrete random variables.

For these two classes of models, the Markov properties impose constraints on the probability distributions which can be interpreted in a nice statistical way. Consider first continuous graphical models.

Let X_V be a vector of p continuous random variables. Assume that

$$P_g = P(X_V|\mu, \Sigma_g) = N_p(\mu, \Sigma_g).$$

Typically, $\mu = 0$. Σ_g is a symmetric and positive definite matrix. The pairwise Markov property is such that a missing edge in the graph is equivalent to a simple constraint on the inverse of Σ, as described by the following.

Definition. Let $K_g = \Sigma_g^{-1}$. Then:

$$(u, v) \notin E \Leftrightarrow k_{u,v} = 0$$

Note that the previous definition implies that a missing edge in the graph g is equivalent to requiring that the corresponding two off-diagonal elements of K are equal to zero. The latter is in turn equivalent to the corresponding partial correlation being equal to zero.

Consider now discrete graphical models. Let X_V be a vector of discrete random variables, whose counts are arranged in a contingency table, $\mathcal{I}$, with cells i.

Let $P_g = \theta_g = (\theta(i), i \in \mathcal{I})$ be a vector of cell probabilities. The markovianity of P with respect to g implies that $(u, v) \notin E \Leftrightarrow$ all loglinear interaction terms involving u and v are equal to zero.

Mixed graphical models, combining both continuous and discrete random variables are also possible, and this is indeed an advantage of the generality of graphical models. In this paper we shall not consider them, mainly because an appropriate Bayesian theory for them has not yet been developed.

What are the advantages of graphical models ? Why an applied statistician should consider them in its bag of tools ? We now list some of the reasons.

- Graphical models can handle very complex stochastic structures in a **unified** manner as for instance, expert systems in artificial intelligence applications, multivariate analysis problems, econometric structural equations, and genetic pedigree analysis. Both discrete and continuous random variables can be dealt with in the same fashion.

- Graphs allow to **read** (write) conditional independences off (in) the graph, helping the discussion on the implementation of statistical methods to analyze real problems.

- Inferences and computations can be carried out **locally**, for small subsets of variables (cliques), with considerable savings in costs and time.

Consider now a complete sample from P, $\underline{X}_V = \underline{x}_V$. An important objective in data mining is model selection, or, in expert systems terminology, structural learning. Statistically this means to choose the most plausible graph g among all candidate graphical structures. In data mining, there is typically little a priori knowledge, so one may want to compare all possible graphs for a given set of random variables.

Our aim will be to select the best graph g and, therefore, the best graphical model P_g, on the basis of the observed sample.

Let $L(\theta_g, g)$ be the likelihood of a graph g, having observed the evidence $\underline{X}_V = \underline{x}_V$.

For model selection, we need to attach a *score* to each considered model. The **classical** (frequentist) score is obtained as

$$S_g = max_{\theta_g} L(\theta_g, g);$$

models are then typically selected via a stepwise selection. Alternatively, one can use penalised likelihood methods, such as AIC or BIC.

The problem with employing classical scores for model selection in data mining is that, as we have already observed, when a large number of variables is considered, stepwise procedures are often very instable. Furthermore, when the purpose is exploratory, that is, there is little subject-matter knowledge on which models are substantially important, as it occurs in data mining, it is advisable to report conclusions from more than one model. Hence a model averaging procedure is needed, and classical methods do not provide an easy solution to this.

The Bayesian approach permit to bypass this problem, as it is based on the comparison of probabilities, which have a natural metric.

The **Bayesian** model score is a (posterior) probability, obtained from:

$$\int_{\theta_g} L(\theta_g, g)\pi(\theta_g)d\theta_g,$$

where $\pi(\theta)$ is the prior distribution of the parameter set. For a non-Bayesian, the prior can be interpreted instrumentally as a weight function and taken as uninformative as possible.

Thus, in a Bayesian approach, each model receives a score which is a coherent probability on the (discrete) space of models. By applying the theorem of total probabilities, it then becomes straightforward to draw inferences, such as posterior means of parameters, as linear combinations of model-specific inferences, with weights described by the model scores. Doing so, the variability of the Bayesian estimates may become large. Although this may be a problem, it is a consequence of the presence of uncertainty on the models.

A problem with the Bayesian approach, that has prevented for long time its applicability, is the need to solve the highly dimensional integrals involved in the derivation of the model scores. This has been made possible, at least approximately, by the upsurge of Markov Chain Monte Carlo methods (MCMC, for a review see e.g. Brooks, 1998).

Recently, Giudici and Green (1999) have proposed a Markov Chain Monte Carlo method based on the Reversible jump algorithm (Green, 1995) for approximate Bayesian model scoring in graphical Gaussian models. Giudici, Green and Tarantola (1999) have extended the approach to consider discrete graphical models. Although here we shall extensively use such methodology, for lack of space we shall only briefly describe it, with reference to the Gaussian case.

Our aim is to construct a Markov chain that will eventually converge to the exact model score. Although the theoretical convergence of the chain will be achieved by construction, it is important to develop diagnostic tools to assess practical convergence, a very difficult task. For a recent proposal see Brooks and Giudici (1999).

The parameter space of interest for a graphical Gaussian model can be described by the triple (g, Σ_g, τ) where g is the (discrete) model indicator, Σ_g is the variance-covariance constrained by the graph and, finally, τ is a set of other random variables contained in the model, such as hyperparameters or missing data. For ease of illustration, here we suppose there is not such τ.

As a prior on Σ_g we consider a Hyper Inverse Wishart distribution (Dawid and Lauritzen, 1993), with hyperparameters Φ and α, here assumed to be known.

The reversible jump MCMC method proposed by Giudici and Green (1999) is able to perform fast Bayesian learning in graphical Gaussian models. The method requires a graph to be decomposable, in order to perform only *local* computations. The algorithm makes use of the incomplete variance-covariance Γ which is such that $\gamma_{ij} = \sigma_{ij}$ iff $(i, j) \in E$ and otherwise unspecified (in fact such latter elements will have to satisfy the non linear constraints imposed by the graph).

4 Applications

4.1 Benchmarking for investment funds

Here the number of possible graphs is equal to $2^{\binom{13}{2}} = 3 \times 10^{23}$; with little prior knowledge selection of only 1 model will clearly lead to an understatement of the variability of the results. Therefore, a Bayesian model selection procedure is clearly preferable.

However, in order to compare results with those obtained with the APT and CAPM models, we first present the estimated regression line that corresponds to the selected graphical Gaussian model, by using a stepwise backward selection procedure that starts from the saturated model on all 13 variables.

$$
\begin{aligned}
& EXTRARETURN = 2.0261 \times BTPind + 0.8783 \times CCTind \\
- \quad & 1.8495 \times Genind - 0.2594 \times ItalianTreasuryBonds \\
+ \quad & 0.1586 \times ForeignBonds - 0.1103 \times ConvertibleBondsandShares \\
- \quad & 0.1417 \times Liquidity,
\end{aligned}
$$

with the last four regressors describing the portfolio composition of the investment fund.

Note that the latter model presents some relevant differences with the CAPM and APT models. This occurs because graphical models correctly take into account also indirect relationships between variables. In particular we have:

- A strong dependence link between the extra return and the BTP index; in fact, if the latter increases the former more than doubles.
- A negative link between the General index and the extra return; the General index is highly correlated with BTP and CCT, indices that mostly influence the extra return and already supply the main information about the correlation with the extra return. Filtering the effect of these two indices we obtain a negative coefficient, different in magnitude with respect to the APT model, and also in sign with respect to the CAPM result.

- A low weight of liquidity and convertible bonds/shares in the ARCA RR portfolio, as underlined by the values of the regression coefficients. In particular, the liquidity, formed by bonds with a short time to run, has a lower return than we obtain from a more long-dated investment.

- A positive effect of the foreign bonds, due to a recent slow increase of the weight of this component in the portfolio of the fund. On the other hand, an inverse relationship between the Italian treasury bonds and the extra return.

The problem with the previous model is that, due to a very high collinearity between the explanatory variables, a classical deviance-based model selection procedure turns out to be quite instable. Furthermore, given the extremely high uncertainty on the association structure, it is important to consider more than one model.

We have thus applied the Bayesian graphical Gaussian data mining procedure described in the previous Section. As a result of $100,000$ iterations of a reversible jump Markov chain, we have obtained a representative graph, containing edges with estimated probability of being present higher than 90%. Such a graph contains 12 edges, and is much more parsimonious that the classical model selected in Section 2, which contained 42 edges. As we now have a more parsimonious model, the predictive power of the selected model is higher.

It is of special interest to evaluate edges from/to the extra-return node. The highest probability of presence is associated to the links with: Italian treasury bonds (0.69), BTP index (0.84), CCT index (1), General index (0.78).

Restricting ourselves to the three most likely benchmarks: BTP, CCT, General index as well as the extra-return we can obtain further insights into the problem.

Figures 1 and 2 show, respectively, the probability distribution (Bayesian model scores) over the graphs and the estimated partial correlations, averaged across all 61 possible decomposable models. In both Figures, node variables are labelled as: $1, 2, 3, 4$, in correspondence with: $Extra - return, BTP, CCT, Generalindex$.

Figure 1 about here

Figure 2 about here

From Figure 1, two structures only account for most of the total probability. The best model shows the extra-return as directly linked only to the BTP and CCT Index, and indirectly with the General Index. The second best model is the saturated model. Note how all explanatory variables are always linked to each other, a consequence of the high collinearity present, especially between the General index and the other two.

Figure 2 gives a clear picture of the additional contribution to the explanation of the extra-return brought by each explanatory variable, when both remaining variables are already present.

Both the CCT and the BTP index residuals are estimated to be negatively correlated with the Extra-return (especially the former) and the General Index is instead positively correlated. This occurs because both BTP and CCT are heavily negatively correlated with the Generale Index, and, therefore, when they are inserted after the latter, they correct negatively its all-purpose positive effect. On the other hand, the General Index net effect is still positive, as it contains further information to determine the extra-return.

Overall, it seems that, if only one benchmark is needed (for interpretational reasons) both the General and the CCT Index seem to be appropriate, with the latter probably preferred in stable (less variable) conditions, and viceversa.

We remark that, if one wanted to obtain similar conclusions, on the basis of the classical analysis, should bear in mind that the results are conditional on the selected model, and, therefore, do depend on the model selection procedure entertained. This may lead to different and possibly biased results, especially when high multicollinearity is present.

Note that the selected final graphical model from the classical procedure is much more parameterised than the Bayesian one and, although this lead to very high levels of variance explanation, cross validation results are slightly worse than those obtained with the simple CAPM model. The Bayesian model that we have obtained here can thus be seen as a useful "compromise".

4.2 Credit Scoring

Consider now the application of our Bayesian discrete graphical models methodology to to the credit scoring dataset. Such analysis will be here described briefly; more details are contained in Giudici, Green and Tarantola (1999).

Figure 3 presents convergence diagnostics for our reversible jump MCMC approach. Because of the sparseness of the considered contingency table, we have run a long simulation, with a burn-in of 50000 and $n = 500000$ subsequent iterations.

Figure 3 about here

Note that the chain reaches stability in correspondence to a simple structure;

the mean number of edges is equal to 11.

On the basis of the previous and others, more formal, convergence diagnostics, we can reliably analyze the output from our reversible jump MCMC simulation.

The high number of possible graphs makes difficult to discriminate between them on the basis of their posterior probabilities. Instead we suggest to build a representative graph, containing edges with estimated probability of being present higher than 90%.

Table 1 presents the estimated edge presence probabilities, needed to build such graph. Each edge is described by a pair of labels, corresponding to the variable indices described in Section 2.

Table 1 about here.

Comparing Table 1 with the results from a classical backward, briefly described in Section2, it turns out that the Bayesian model is more parsimonious. Credit reliability is conditionally independent on the variables which were also previously such, however there is one further independence, with marital status.

In addition, the Bayesian model allows to estimate, as done in Table 1, the probability that a particular link between two variables is supported by the data at hand, regardless of the chosen model. This is an example of a model averaged inference which can be obtained following a Bayesian approach. Such inferences do not require to choose a particular structure, and turn out to be rather stable, both with respect to prior specification and in terms of the MCMC output.

5 Conclusions

Graphical models do indeed represent a powerful methodology for data mining. The decomposable undirected models considered here allow, as we have seen, to extract very valuable information from the available datasets, in the form of conditional dependence relationships which were not known a priori.

By means of the reversible jump MCMC method presented, we have also been able to attach to each alternative association structure a score, which is simply interpretable, being a (posterior) probability. In data mining contexts, it is quite difficult to restrict attention to a small subset of models and, thus, the presented methodology, which allows an efficient search in the space of all models, may also be quite valuable.

We remark that the MCMC Bayesian model selection methodology presented here has been introduced very recently. In this paper we have applied it to two real and challenging datasets, and have shown that new insightful information on the association structure can be extracted, in a way that is also simply understood and communicated.

We believe that more applied research should be done in the area, exploiting the full potentials of Bayesian graphical models. Furthermore, more methodological research is needed. For instance, mixed models should be analyzed. Also more complicated structures of graphs, including directed edges and latent variables, should be brought into this framework.

Another open issue of research is to find computational to accelerate model selection procedures on a large number of variables. In particular, methods that speed up the practical convergence of the reversible jump algorithm would be a very important asset to make it suitable for the analysis of data mining problems.

6 References

Brooks, S P. (1998). Markov Chain Monte Carlo Method and its Application. *The Statistician*, **47**, 69-100.

Brooks, S. P., and Giudici, P. (1999). MCMC Convergence Assessment via Two-way ANOVA. *Journal of Computational and Graphical Statistics*, to appear.

Dawid, A.P. and Lauritzen, S.L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics* **21**, 1272–317.

Elton, J. and Gruber, M.J. (1995). *Modern portfolio theory and investment analysis.* Wiley, Chichester.

Fahrmeir, L. and Hamerle (1984). *Multivariate statistische Verfahren.* De Gruyter, Berlin.

Francis, J.C. (1991). *Investments analysis and management.* Mc Graw-Hill, New York.

Giudici, P. and Green, P. J. (1999). Decomposable Graphical Gaussian Model Determination. *Biometrika*, **86**, 785-801.

Giudici, P., Green, P.J., and Tarantola, C. (1999). Efficient Model Determination for Discrete Graphical Models. Technical Report, Athens University of Economics and Business.

Green, P. J. (1995). Reversible jump MCMC Computation and Bayesian Model Determination. *Biometrika*, **82** 711–732.

Hand, D.J. and Henley, W.E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of The Royal Statistical Association A*, **160**, 523-541.

Hand, D.J., McConway, K.J. and Stanghellini, E. (1997). Graphical Models of Applicants for Credit. *IMA Journal of Mathematics Applied in Business and Industry*, **8**, 143-155.

Lauritzen, S.L. (1996). *Graphical models.* Oxford University Press, Oxford.

Madigan, D. and Raftery, A.E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of The American Statistical Association*, **89**, 1535-46.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics.* Wiley: New York.

Table 1: Credit scoring: estimated probability of an edge being present

Edge	**1/100**	**Edge**	**1/100**	**Edge**	**1/100**
[2,1]	1.000000	[6,3]	0.000270	[8,4]	0.004520
[3,1]	0.000910	[6,4]	0.000155	[8,5]	0.651470
[3,2]	0.000000	[6,5]	0.014205	**[8,6]**	1.000000
[4,1]	0.000305	[7,1]	0.022965	[8,7]	0.005060
[4,2]	0.000380	[7,2]	0.004690	[9,1]	0.093475
[4,3]	1.000000	[7,3]	0.009405	[9,2]	0.037830
[5,1]	0.018675	[7,4]	0.034695	[9,3]	0.892105
[5,2]	0.002710	[7,5]	0.288675	**[9,4]**	0.999835
[5,3]	0.010265	[7,6]	0.005105	**[9,5]**	1.000000
[5,4]	0.229645	**[8,1]**	0.917005	[9,6]	0.010960
[6,1]	0.121920	[8,2]	0.049140	[9,7]	0.766750
[6,2]	0.003020	[8,3]	0.005115	**[9,8]**	0.999805

Figure 1: Benchmarking : estimated posterior model probabilities over the restricted four vertices dataset.

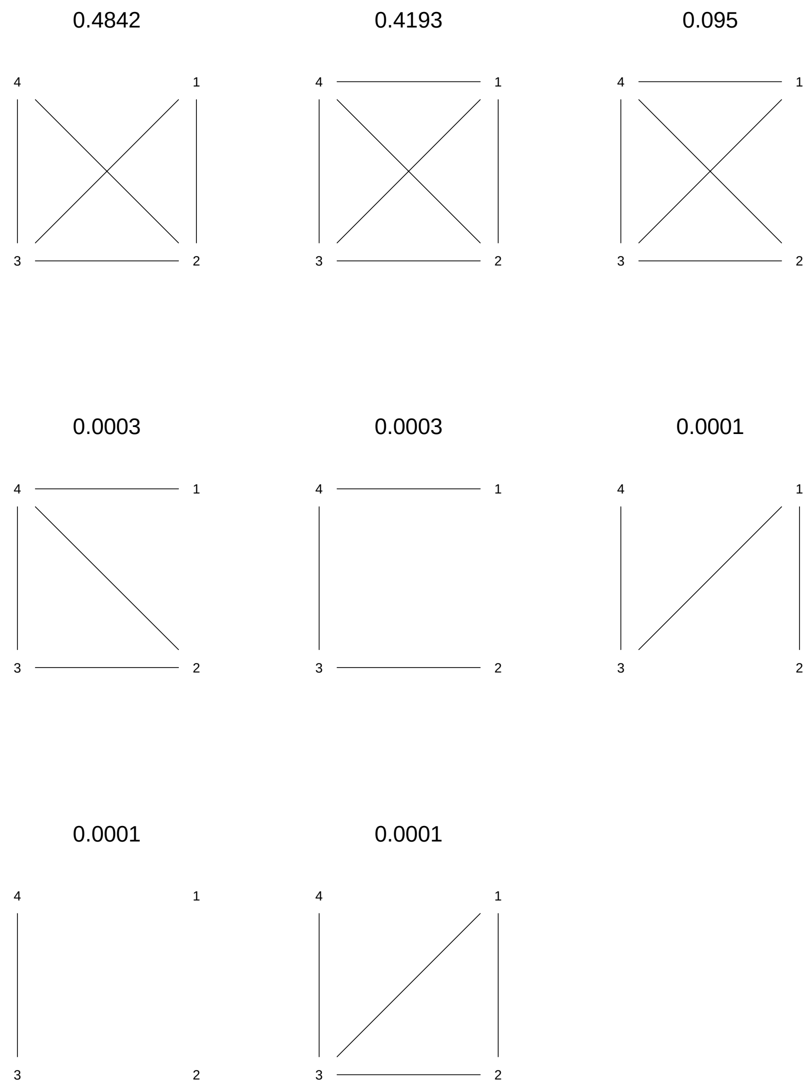

Figure 2: Benchmarking : estimated posterior distribution of the partial correlations for the restricted dataset.

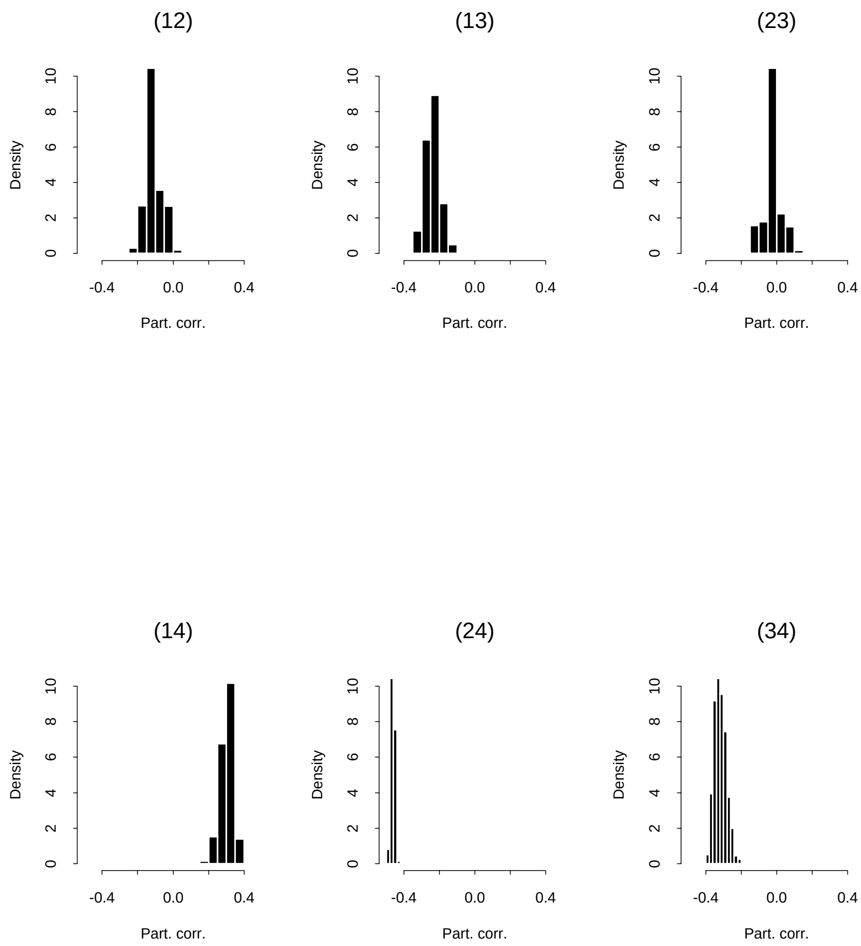

Figure 3: Credit scoring : diagnostics on the number of edges present.